

Clinical Vocabularies - Their Potential as Classifications

Charlie D.G. Stuart-Buttle M.B.,B.S., John S.A. Ashley F.R.C.P., F.F.C.M.,
Mike O'Neil M.B.,B.S.

NHS Centre for Coding and Classification, Loughborough, England

Background. There has been considerable debate about the relative merits of vocabularies and classifications^{1,2}. We believe that the debate is not yet over. Here we explore the nature of vocabularies and classifications and the potential issues involved in vocabularies, in particular Read Version 3³, being used to provide comparable statistical data.

The nature of vocabularies. Modern clinical vocabularies have a number of important features that differentiate them from classifications. These are responsiveness to change, concepts described in the detail required to record everyday clinical care, multiple classification, and a flexible hierarchy.

The nature of classifications. The features of statistical classifications that are relevant to this debate are stability over time, exhaustiveness, mutual exclusivity of categories and a single view of the world. Most classifications are designed to provide output and not for use as input devices.

Aggregation using a vocabulary. Searches of clinical databases are broadly of two essential sorts: retrieval of all cases of a single clinical entity; and aggregations of data. It is aggregation that poses the greatest problems to dynamic vocabularies.

The crucial question is whether aggregation for particular purposes can occur using the class structure in the original vocabulary or whether there is a need for transformation to another form, often a statistical classification. One needs to examine who uses statistical data and for what. A new approach is probably called for. It may be that some of the data traditionally gathered will not be derivable using the vocabulary-based approach. We can then ask whether it is really required.

Stability of vocabularies and aggregation. Data comparison requires stability over time. Further study is needed of the statistical significance of the effect of change in dynamic vocabularies on the analysis of aggregated data over time.

A stable thesaurus. There are a number of possible

ways to combine stability with the dynamic nature of a vocabulary such as Read Version 3. Maximising accuracy and completeness in the semantic definition and hierarchical placement of concepts will provide consistency of representation, allowing analysis using more static aspects such as site. A fixed classification hierarchy might be constructed by nominating a single parent-child link as the classification link where multiple parents exist. Re-running searches on previous years' data using the latest release of Version 3 with any changes it may include may be a possible approach.

Aggregation using maps. An alternative is to collect data using a vocabulary and then, via maps, to aggregate using classifications. This itself is far from simple as other factors come into play. In Read Version 3 the nature of maps, their stability across updates, the proportion of one-to-one maps, the effect of qualifiers and the rules of the classifications themselves all have an influence, as does inter-professional variability in using such tools.

Conclusion. It is our view that data should ideally be both collected and aggregated using a controlled clinical vocabulary unless a better tool is available. Much remains to be explored in this area. In the mean time using a combination of fixed classifications with dynamic clinical vocabularies and ensuring that the links between them are clear and accurate seems the sensible way forward.

References

1. Côte RA. Editorial: Ending the classification versus nomenclature controversy. *Medical Informatics* 8, 1993, 1-4.
2. Ingenerf J, Giere W. Concept-oriented Standardization and Statistics-oriented Classification: Continuing the Classification versus Nomenclature Controversy. *Proceedings of the IMIA WG6 Conference*. Jacksonville, 1995 (in press).
3. O'Neil MJ, Payne C, Read JD. Read Codes Version 3: A User Led Terminology. *Meth Inform Med* 1995; 34: 187-92.